## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

**Patent Application**

| | |
|---|---|
| Applicant(s): | Chen et al. |
| Docket No.: | YOR919990172US1 |
| Serial No.: | 09/345,238 |
| Filing Date: | June 30, 1999 |
| Group: | 2626 |
| Examiner: | Qi Han |
| Title: | Method and Apparatus for Tracking Speakers in an Audio Stream |

## REPLY BRIEF

Mail Stop Appeal Brief – Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Appellants hereby reply to the Examiner's Answer, mailed July 21, 2008 (referred to hereinafter as "the Examiner's Answer"), in an Appeal of the final rejection of claims 1-35 in the above-identified patent application.

## REAL PARTY IN INTEREST

A statement identifying the real party in interest is contained in Appellants' Appeal Brief.

## RELATED APPEALS AND INTERFERENCES

A statement identifying related appeals is contained in Appellants' Appeal Brief.

## STATUS OF CLAIMS

A statement identifying the status of the claims is contained in Appellants' Appeal Brief.

## STATUS OF AMENDMENTS

A statement identifying the status of the amendments is contained in Appellants' Appeal Brief.

## SUMMARY OF CLAIMED SUBJECT MATTER

A Summary of the Invention is contained in Appellants' Appeal Brief.

## STATEMENT OF GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A statement identifying the grounds of rejection to be reviewed on appeal is contained in Appellants' Appeal Brief.

## CLAIMS APPEALED

A copy of the appealed claims is contained in an Appendix of Appellants' Appeal Brief.

## ARGUMENT

In the Examiner's Answer, the Examiner asserts that claim 1 includes only two steps and that this makes the claimed limitation of "substantially concurrently" to be less clear and lack patentable weight. The Examiner asserts that Appellants' arguments based on this limitation appear to try to exclude the Examiner's interpretation that the claim includes that "clustering and segmenting are performed sequentially" and try to cover the "claimed two steps that, in fact, are performed sequentially in light of specification." The Examiner asserts that these arguments are either contrary to themselves, or imply that the limitation of "substantially concurrently" does not exclude the Examiner's interpretation.

Appellants note that "substantially" is defined as "to a great extent or degree" and that "concurrently" is defined as "overlapping in duration." (See, dictionary.com) Thus, "substantially concurrently" means "overlapping in duration to a great extent or degree." As can be seen, contrary to the Examiner's assertion, "substantially concurrently" eliminates the Examiner's interpretation that the "clustering and segmenting are performed sequentially" (since "sequence" implies no overlap), but includes the cases *where the clustering and segmenting are performed <u>concurrently</u>* and the case *where the clustering and segmenting are performed* *concurrently* *to a great extent or degree*.

Regarding the Examiner's assertion that a feature (the loop of FIG. 2) upon which applicant relies is not recited in the rejected claims, Appellants note that the loop of FIG. 2 was presented in the arguments as support for the limitations at issue (segmentation and clustering are performed substantially concurrently or during the same pass through the audio source). The cited limitations *are recited in the rejected claims*.

The Examiner further asserts that, since the limitation "substantially concurrently" is not specifically defined and linked to the loop in the specification, it has much broader scope than the argued "loop" mechanism.

As noted above, the term "substantially concurrently" is well understood by a person of ordinary skill in the art and is linked to the loop of FIG. 2 by the fact that the segmentation and clustering of FIG. 2 can be performed *concurrently* *to a great extent or degree*.

The Examiner asserts that the cited claims recite two steps, that FIG. 2 shows that the corresponding steps are performed sequentially, and that this provides evidence that the rejection based on the Examiner's interpretation is proper.

Appellants note that the Examiner's interpretation is improper since it ignores the limitation that the clustering and segmentation are performed substantially concurrently or during the same pass (as required by the rejected claims). Furthermore, Appellants note that the loop of FIG. 2 provides additional support for this limitation and that the limitation should therefore be given patentable weight.

The Examiner asserts that appellant's argument that "segmentation may be performed both before and after clustering" is meaningless because, even in the local process (i.e., within a loop), segmenting (is) always before clustering and that there is no evidence in the application/arguments that shows any useful application to perform segmentation after clustering.

Contrary to the Examiner's assertion, the step or process of segmenting may continue after clustering has begun. Moreover, clustering may be *temporarily halted while segmentation continues*. Thus, segmentation may occur before, during, or *after* clustering (e.g., after clustering has been temporarily halted).

Regarding the Examiner's assertion that appellant's arguments on page 8, paragraph 6, to page 10, paragraph 1, are extra features of what the prior art is doing, Appellants note that these arguments *provide additional evidence to support the fact that the features and limitations relied upon by the Appellants are not disclosed or suggested by the prior art.*

Appeal Brief Arguments

Independent Claims 1, 16, 23, and 30-35

Independent claims 1, 16, 23 and 30-35 are rejected under 35 U.S.C. §102(b) as being anticipated by Chen et al.

In the Office Action dated August 27, 2002, the Examiner asserted that Chen discloses speaker, environment and channel change detection and clustering via the Bayesian Information Criterion for segmenting the audio stream into homogeneous regions according to speaker identity, environmental condition and channel condition and clustering speech segments into homogeneous clusters according to speaker identity, environmental condition and channel (citing page 1, paragraph 2) which reads on the claimed "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently  with said identifying step."

In the Response to Office Action dated December 26, 2002, Appellants submitted that, while Chen discloses segmenting an audio stream into homogeneous regions and clustering speech segments into homogeneous clusters, the audio stream is *first* segmented and *then* clustered. Appellants noted, as further evidence that the clustering in Chen is performed only after the audio stream has been segmented, that Section 4.1 indicates that each segment is compared to all other segments before clustering is finalized. In addition, Section 4.2, first paragraph indicates that the data set  consists of an audio file that has been "hand-segmented into 824 short segments."

In the Office Action dated March 7, 2003, the Examiner notes that the prior art cites that "our segmentation algorithm can successfully detect acoustic changes" (Chen: abstract) and that "we first examine whether our detected change points were true." (Chen: Section 3.3, paragraph 3.) The Examiner asserts that this suggests that Chen not only employs its own segmenting mechanism, but is also capable of combining segmentation with clustering "substantially concurrently."

The Examiner also asserts that Chen suggests that the clustering does not need completely segmented data, such that a clustering process may be combined with a segmenting process together substantially concurrently, since Chen discloses that "it is also clear that our criterion can be applied to top-down methods." (Chen: Section 4.1, paragraph 4.)

The Examiner further asserts that a clustering step can be inserted in the segmentation loop, in Chen, Section 3.2, paragraph 1, and that Chen is capable of combining segmentation and clustering since the segmentation and clustering algorithms are based on the BIC algorithm and since equations (2), (3), and (8) have no limitation for combining segmentation and clustering.

Appellants acknowledge that Chen employs its own segmenting mechanism, but find no indication of or suggestion to perform segmentation and clustering "substantially concurrently" in the cited text. Appellants note that the Examiner asserts that Chen is *capable of* this, but does not assert that Chen suggests or discloses combining segmentation with clustering substantially concurrently.

Appellants also note that, in the top-down method, a hypothesis is made regarding the number of clusters. Then, a test is made to determine if the number of clusters hypothesized actually "fits" the data. Alternatively, in the bottom-up method, the number of clusters is determined from the data. Thus, the capability to utilize a top-down method does not suggest that segmentation is performed substantially concurrently with the clustering process.

Regarding the final assertion made by the Examiner, Appellants also note that, whether or not Chen is *capable* of combining segmentation and clustering, there is no disclosure or suggestion to do so.

Thus, Chen does not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Response to Examiner's Answer dated December 17, 2003

In the Examiner's Answer dated December 17, 2003, the Examiner states that it is believed that the limitation "substantially concurrently" has no patentable weight, because the

Applicant does not have any clear definition and/or description in the claim or in the specification about this limitation, and does not give any conditions to apply this limitation. The Examiner also asserts that the prior art explicitly and/or implicitly discloses all the limitations regarding claim 1, including the limitation of "substantially concurrently," based on the interpretation of the claim language and the understanding (of the) prior art teachings. In particular, the Examiner asserts that the performance of the two steps (segmentation and clustering) may be associated with many time related factors, including computing speed, simple rate, and total stream size.

The Examiner further asserts that the fact that the clustering in Chen is performed only after the audio stream has been segmented and that each segment is compared to all other segments before clustering is finalized is not relevant to claim 1 since claim 1 does not recite these limitations.

The Examiner also notes that one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references.

Regarding the Examiner's assertion that the limitation "substantially concurrently" has no patentable weight, Appellants note that the word "substantially" has a well known and well understood definition in claim language. Its meaning is sufficiently clear in the teachings of the specification such that a person of ordinary skill in the art would understand the limitation without the need to apply conditions.

Regarding the Examiner's assertion that the prior art explicitly and/or implicitly discloses all the limitations regarding claim 1, including the limitation of "substantially concurrently," based on the interpretation of the claim language and the understanding (of the) prior art teachings, Appellants note that the broad interpretations made by the Examiner are ***not consistent*** with the specification and are ***not consistent*** with the interpretation of the specification that a person of ordinary skill in the art would make. As disclosed on page 2 (lines 16-26) of the original specification, "the present invention concurrently *segments an audio file and clusters the segments* corresponding to the same speaker." Thus, the term "substantially concurrent" is related to the *parallel execution* of the segmentation and clustering steps. See, also, FIG. 2.

More specifically, Appellants note that these limitations are clearly captured in claim 1, which recites the limitations of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step. Claim 1 requires the clustering of homogeneous segments *substantially concurrently* with said identifying step. Chen, therefore, actually teaches away from the present invention by teaching that the clustering is performed only after the audio stream has been segmented. Thus, contrary to the Examiner's assertion, the limitations cited by the Examiner in reference to Chen are ***clearly relevant*** to the consideration of claim 1.

Appellants also note that the references were not attacked individually, but were reviewed to demonstrate that ***none*** of the references contain the cited limitation required by the claims of the present invention and that, therefore, the prior art does not pose a bar to patentability.

Response to Final Office Action of December 18, 2007

The Examiner asserts that the claimed "clustering homogeneous segments...substantially concurrently with said identifying step (segmentation)" does not exclude the situation that "the clustering is performed only after the audio stream has been segmented."

Appellants note that the Examiner is including the case where the *clustering and segmenting are performed sequentially.* This is contrary to the claim requirement that the *clustering and segmenting are performed "substantially concurrently."* The term "substantially concurrently" should be given patentable weight. In the cited example, however, the clustering and segmentation are performed sequentially; there is absolutely no degree to which the clustering and segmentation are performed "substantially concurrently." Thus, the Examiner's interpretation of the cited claims is *not a reasonable interpretation.*

Moreover, the *loop* illustrated in FIG. 2 demonstrates that the segmentation and clustering are performed substantially concurrently, as segmentation may be performed both *before and after clustering*.

The Examiner further asserts that "Chen's disclosure satisfies the claimed limitation under at least this minimum condition/assumption, because Chen recites 'comparing two models, one models the data as two Gaussian(s); the other models the data as just one Gaussian' to detect the changing point for segmentation (Chen, Section 3.1, page 4), such that at least two data groups (segments) are segmented (before) for clustering speakers (Chen, Section 4, page 8)."

Appellants note that Chen describes a Gaussian distribution with means and variances. Chen assumes that the means of all the signals, s1 ... sk, can be computed. This is only feasible if all the data required to compute the mean is available. For example, imagine there is a pipe conveying water from point A to point B. The observer at point B does not know how water will come over. (This is analogous to a radio signal or video stream.) In order to compute the mean volume of water emanating from a pipe per unit time, all the water can be collected into a container, the time required for the pipe to go dry can be measured, and the volume of water collected can be divided by the time required to collect it. This, in effect, is Chen's approach. In one aspect of the present invention, a running mean is used - that is, as and when the data arrives its statistics are computed. In terms of the cited example, the volume of water arriving every second (or some fixed multiple) is measured and running statistics are maintained. The result is therefore usable from the time the water starts emanating from the pipe.

In one embodiment of the present invention, the segments are automatically computed, where each segment is a speaker turn in a conversation. By gathering together all "similar" segments into a cluster, all of these speaker turns are recognized to correspond to one individual. This is done when the person finishes speaking his or her turn. In a roundtable of speakers, speech by the same speaker is able to be segmented and clustered as and when it occurs, versus after the entire roundtable has ended.

Chen, alternatively, outlines two clustering approaches in Sections 4 and 4.1 . In the clustering approach of Section 4, the audio is first broken up into segments using the BIC criterion. The clustering begins after the entire audio has been broken into segments. In the real world, when dealing with real-time video or audio streams, the "entire" audio can be acquired, for example, only after one hits the 'stop' button on the recorder. After breaking the audio down into individual segments, Chen collects them into clusters. The number of clusters is open to begin with as is the cluster membership. Chen combines audio segments in different combinations in order to arrive at a globally optimized set of clusters as defined by the BIC criterion. As is stated in the first paragraph of Section 4.1, the process is very computationally expensive.

In the second "greedy" clustering approach (Section 4.1), Chen's starting point is the same as the above. It is a set of individual audio segments realized "after" the entire audio stream has been broken up into segments. There is no simultaneous segmentation and clustering. Clustering follows segmentation; only the clustering algorithm is slightly optimized over the first technique. Chen states in the first sentence of section 4.1 that the process of clustering works by merging nearest nodes. Chen can do this only after he has access to "all" the nodes (each node corresponds to a segment). In line 3, paragraph 2 of Section 4.1, Chen teaches, "let S = {s1, s2, ..., sk} be the current set of nodes ..." Here, Chen tacitly states that there is access to all the segments labeled s1 through sk, where k is the total number of segments (nodes), i.e., there is access to all the audio that is desired to be analyzed. In any real-time application, such as streaming audio or streaming video, there is access to all that has transpired thus far. Thus, segmentation and clustering can only be done based on past events.

Additional Cited References

Kleider et al. was also cited by the Examiner in rejecting claim 15 for its disclosure that the information of the speaker model data may include a speaker name. Appellants note that the inventors listed in United States Patent Number 5,157,763 (referred to by the Examiner in the Final Ofiice Action) are not Kleider et al. Appellants did find, however, United States Patent Number 5,930,748 in the Notice of References Cited and respond to that reference below.

Appellants note that Kleider et al. is directed to a "method of identifying an individual from a predetermined set of individuals using a speech sample spoken by the individual. The speech sample comprises a plurality of spoken utterance, and each individual of the set has predetermined speaker model data." Cited, Summary of the Invention. Kleider et al. do not address the issue of segmenting speech.

Thus, Kleider et al. do not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Conclusion

The rejections of the cited claims under sections 102 and 103 in view of Chen, Kleider et al. or well known prior art, alone or in any combination, are therefore believed to be improper and should be withdrawn. The remaining rejected dependent claims are believed allowable for at least the reasons identified above with respect to the independent claims.

The attention of the Examiner and the Appeal Board to this matter is appreciated.

Respectfully,

Date: September 17, 2008

Kevin M. Mason
Attorney for Applicant(s)
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560

APPENDIX

1. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries in said audio source; and

clustering homogeneous segments from said audio source substantially concurrently with said identifying step.

2. The method of claim 1, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

3. The method of claim 2, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

4. The method of claim 2, wherein a given sample, i, in said audio source is likely to be segment boundary if the following expression is negative:

$$\Delta BIC_i = -\frac{n}{2}\log|\Sigma_w| + \frac{i}{2}\log|\Sigma_f| + \frac{n-i}{2}\log|\Sigma_s| + \frac{1}{2}\lambda\,(d + \frac{d(d+1)}{2})\log n$$

where $|\Sigma_w|$ is the determinant of the covariance of the window of all n samples, $|\Sigma_f|$ is the determinant of the covariance of the first subdivision of the window, and $|\Sigma_s|$ is the determinant of the covariance of the second subdivision of the window.

5. The method of claim 1, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

6. The method of claim 5, wherein said window size, n, is increased in a relatively slow manner when the window size is small and increases in a faster manner when the window size is larger.

7. The method of claim 5, wherein said window size, n, is initialized to a minimum value after a segment boundary is detected.

8. The method of claim 2, wherein said BIC model selection test is not performed at the border of each window of samples.

9. The method of claim 2, wherein said BIC model selection test is not performed when the window size, n, exceeds a predefined threshold.

10. The method of claim 1, wherein said clustering step is performed using a BIC model-selection criterion.

11. The method of claim 10, wherein a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

12. The method of claim 11, further comprising the step of merging said two clusters if a difference in BIC values for each of said models is positive.

13. The method of claim 1, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

14. The method of claim 1, further comprising the step of assigning a cluster identifier to each of said clusters.

15. The method of claim 1, further comprising the step of processing said audio source with a speaker identification engine to assign a speaker name to each of said clusters.

16. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries in said audio source; and

clustering segments from said audio source corresponding to the same speaker substantially concurrently with said identifying step.

17. The method of claim 16, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

18. The method of claim 17, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

19. The method of claim 16, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

20. The method of claim 17, wherein said BIC model selection test is not performed where the detection of a boundary is unlikely to occur.

21. The method of claim 16, wherein said clustering step is performed using a BIC model-selection criterion, where a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

22. The method of claim 16, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

23. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries during a pass through said audio source; and clustering segments from said audio source corresponding to the same speaker during said same pass through said audio source.

24. The method of claim 23, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

25. The method of claim 24, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

26. The method of claim 23, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

27. The method of claim 24, wherein said BIC model selection test is not performed where the detection of a boundary is unlikely to occur.

28. The method of claim 23, wherein said clustering step is performed using a BIC model-selection criterion, where a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

29. The method of claim 23, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

30. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries in said audio source; and

cluster homogeneous segments from said audio source substantially concurrently with said identification of segment boundaries.

31. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries in said audio source; and

a step to cluster homogeneous segments from said audio source substantially concurrently with said identification of segment boundaries.

32. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries in said audio source; and

cluster segments from said audio source corresponding to the same speaker substantially concurrently with said identification of segment boundaries.

33. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries in said audio source; and

a step to cluster segments from said audio source corresponding to the same speaker substantially concurrently with said identification of segment boundaries.

34. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries during a pass through said audio source; and

cluster segments from said audio source corresponding to the same speaker during said same pass through said audio source.

35. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries during a pass through said audio source; and

a step to cluster segments from said audio source corresponding to the same speaker during said same pass through said audio source.

## EVIDENCE APPENDIX

There is no evidence submitted pursuant to § 1.130, 1.131, or 1.132 or entered by the Examiner and relied upon by appellant.

## RELATED PROCEEDINGS APPENDIX

There are no known decisions rendered by a court or the Board in any proceeding identified pursuant to paragraph (c)(1)(ii) of 37 CFR 41.37.